The Nevada Adequate Yearly Progress Technical Manual Revision 3

June 7, 2006

Prepared by:

Paul M. La Marca, Ph.D.
Assistant Deputy Superintendent
Nevada Department of Education

For additional copies, please visit our website at www.doe.nv.gov or contact the APAC office at (775) 687-9186. Please submit questions to Paul La Marca at plamarca@doe.nv.gov or at (775) 687-9180.

The Nevada Adequate Yearly Progress Determination Process

On January 9th, 2002 President Bush signed into law the Federal No Child Left Behind Act (NCLB—HR 1). This reauthorization of the federal Elementary and Secondary Education Act and its sweeping reforms has impact on every state including Nevada. In response to the new federal law, the Nevada Legislature significantly revised its own accountability statutes through passage of Senate Bill 1 in the 19th Special Session (June, 2003). This was a necessity given significant differences between existing state and federal statute. At the heart of both the federal and revised state statutes is a conservative school, school district, and state accountability model working under the auspice of guaranteeing all students the opportunity for and access to a challenging and meaningful educational experience. Toward this end and on an annual basis, schools, school districts and the state as a whole are judged against a set of adequate yearly progress (AYP) criteria. The judgment of success is based largely on performance on assessments aligned to state content standards administered on an annual basis, and by attending specifically to the performance of disparate subgroups of students.

Much has transpired since the passage of the Act. Nevada has complied with the requirements of the law and has made annual school and district determinations in each of the last three school years. The next set of determinations will occur in the summer of 2006 based largely on test results from the spring of the 2005-06 school year. Additionally, the federal government has in specific instances provided states with flexibility on how they apply AYP requirements. Moreover, the State of Nevada has requested changes to its AYP model in each of the last three school years. Changes that have impacted the determination process and pending changes that would create a technical difference in making determinations will be addressed in this document.

The following is an overview of the Nevada AYP Determination Process as it stands now. The Nevada AYP model is likely to see additional changes in years to come. When those changes introduce significant technical differences, a revision to this bulletin will be provided. In this manual, the Department focuses on the following elements:

- The Assessment system
- The AYP Determination Process
- Critical Timelines
- Reporting Issues
- Special schools and circumstances

It is important to note that the AYP determination process applies equally to schools, school districts, and the state. In this document, schools are the descriptive unit of analysis. The same general rules are applied to school districts and the state. Note will be made when differences in the application to districts or the state do exist.

The Assessment System

The foundation for the NCLB accountability system is made up of state content and achievement standards and large scale assessments designed to measure the standards. NCLB expanded previous requirements regarding the development of state standards in English Language Arts and Math by requiring states to develop content and achievement standards in science.

Similarly, NCLB expanded assessment requirements from previous legislation by requiring states to develop and implement tests in grades 3 through 8 and in at least one grade at the high school level in English Language Arts and Math. In compliance with the law, reading and math tests are being administered live in grades 3 through 8 this school year (2005-06)[1]. Reading and math tests in grades 3 and 5 have been administered since 2000-01, and 8th grade tests were added in 2003-04. Additionally, by 2007-08, states must develop and implement science tests to be administered in at least one grade in three separate grade ranges (3-5, 6-8, 9-12). Nevada currently administers science tests in grades 5 and 8 and is developing a high school exit examination in science. This latter test will be administered live for the first time in the 2007-08 school year. Science tests are not included in making AYP determinations.

A couple of other assessment related statutory requirements are pertinent. Assessments to be used in the AYP determination process must directly align to state content standards. This led to Nevada's use of its existing criterion-referenced tests (CRT) and the expansion of that program to meet federal requirements. Additionally, as a result of the alignment requirement, Nevada no longer uses norm-referenced tests (NRT) when making formal judgments about schools. The state does require the use of NRTs in grade 4, 7, and 10. Results from these tests are reported in annual accountability reports and are used to validate CRT performance.

The alignment condition also has a direct relationship to the federal requirement that states employ multiple measures in assessing student achievement. The use of multiple measures is expected to enhance both the reliability and validity of the measurement process. It does this by enabling an expansion of the content assessed and allowing the convergence of findings from disparate measurement methods. The relationship to alignment is evidenced when we draw a distinction between breadth of content coverage and depth of content coverage.

Breadth of coverage is achieved by sampling widely and representatively from the stated or prescribed content. Depth of coverage refers to the degree by which the cognitive demands elicited by the assessment tasks match the cognitive demands prescribed by the content standards. For example, we expect students to be able to write persuasive and informational essays that clearly articulate ideas, are well organized, employ appropriate voice, and are

[1] As indicated in Nevada's accountability work plan, the timing associated with setting achievement standards on the grade 4, 6, and 7 tests prevents their inclusion in the AYP determination process for the 2005-06 school year. However, the Department will be issuing in early fall of 2006-07, for information only, an analysis of what would have been the AYP result if the grade 4, 6, and 7 tests had been included in the AYP analysis.

conventionally sound. For measurement purposes, we have available to us multiple choice questions that correspond to idea articulation, organization, voice, and conventions. Student responses to these questions may tell us something about writing ability. However, there may be alternate methods available that provide a more direct measurement of writing skill and that more directly tap the cognitive demands prescribed by the standards. In other words, having students respond openly to a writing prompt may entail a very different cognitive process than having students respond to an editing activity.

Understanding this distinction and the federal requirement, CRTs used in the AYP process have multiple question formats including multiple choice items and short open-response items. The state continues to administer performance-based writing tests.

Tests used in the AYP process must also align with state achievement standards[2]. This means that the tests must enable a distinction to be drawn between students that differ in their demonstration of achievement. Minimally, tests must allow the classification of students into three achievement levels (e.g. basic [below proficient], proficient, and advanced). The federal government allowed flexibility in the labeling of achievement levels and in the number of achievement levels as long as the above distinctions were met. In Nevada four achievement levels are used, with two levels identifying performance that is below meeting standard or "proficient".

The key point to be made is that each assessment used in the AYP process must yield information that can be categorized using the achievement level distinctions. As will be discussed in more detail, this similarity among assessments included when determining AYP allows the results to be combined when making AYP determinations.

Table 1. Crosswalk of Nevada and Federal Achievement Level Categories

Nevada Achievement Levels	Federal Achievement Levels
Developing/Emergent	
Approaching Standard	Basic
Meets Standard	Proficient
Exceeds Standard	Advanced

States have the responsibility to determine what level of performance (cut-scores) on its tests are indicative of proficiency or of meeting the state's content standards or expectations for student knowledge and skill attainment. States must employ objective methodologies that rely on the professional judgment of educators in making these decisions.

States may also use an alternate assessment for students with severe cognitive disabilities. This is predicated on strict eligibility criteria and must be clearly stated in a student's individualized education program (IEP). The federal government places a cap of 1% of the total student population that can be counted as proficient based on the use of alternate assessments.

[2] In Nevada, content standards describe what a student should know and be able to do by the end of a particular grade level. Achievement standards (also known as performance standards) provide a description of what students must demonstrate to be classified along an achievement continuum.

Table 2 includes a summary of the tests used in Nevada's AYP determination process and demonstrates Nevada's compliance with the assessment requirements of the NCLB Act.

Table 2. Current Nevada Tests Included in Determining AYP

	Tests Administered in 2005-2006 Included in the AYP Process
Grade 3	CRT—Reading, Math NASAA—ELA, Math
Grade 4	CRT—Reading, Math NASAA—ELA, Math
Grade 5	CRT—Reading, Math Performance—Writing NASAA—ELA, Math
Grade 6	CRT—Reading, Math NASAA—ELA, Math
Grade 7	CRT—Reading, Math NASAA—ELA, Math
Grade 8	CRT—Reading, Math Performance—Writing NASAA—ELA, Math
High School	HSPE—Reading, Math, Writing NASAA—ELA, Math
CRT = Criterion-referenced tests HSPE = High School Proficiency Examination NASAA = Nevada Alternate Scales of Academic Achievement ELA = English Language Arts (includes reading & writing)	

Nevada has used an adapted Bookmark method in establishing most of its test cut-scores. Cut-scores for the grade 4, 6, and 7 tests will be set in the summer of 2006. In setting those cut points, steps will be taken to ensure their consistency with cut points at other grades. Table 3 provides a summary of test cut scores that correspond to the achievement levels used in the AYP determination process.

In compliance with the NCLB Act, the Nevada accountability plan allows school districts to opt for use of alternate content-based assessments for limited English proficient students who have been in the United States for less than a 3-year period. On a case-by-case basis, districts may choose to extend the use of alternate assessments for two additional years. The use of alternate assessments must be approved by the Nevada Department of Education prior to use. Approval is based on a review of technical characteristics of the assessment. In short, the alternate assessments must be aligned to the state's content and achievement standards and must yield reliable and interpretable performance information. It is important to note that this section refers to subject area tests used in the AYP determination process and not to tests required by the federal government as part of Title III.

Table 3. Achievement Level Cut-Scores

Grade	Test[a]	Subject	Developing/ Emergent	Approaches Standard	Meets Standard	Exceeds Standard
Grade 3	CRT	Reading	100-199	200-299	300-355	356-500
		Math	100-199	200-299	300-354	355-500
Grade 5	CRT	Reading	100-199	200-299	300-384	385-500
		Math	100-199	200-299	300-380	381-500
	Performance	Writing	0-7.5	8-11.5	12-15.5	16-20
Grade 8	CRT	Reading	100-199	200-299	300	385-500
		Math	100-199	200-299	300	385-500
	Performance	Writing	0-7.5	8-11.5	12-15.5	16-20
Grades 10 & 11	HSPE	Reading	TBD	TBD	251	TBD
		Math	TBD	TBD	290	TBD
		Writing	TBD	TBD	7	TBD

[a] Achievement level cut points for the NASAA tests are not shown.

Section Summary

The requirements for standards and assessments within the NCLB Act lay the foundation for the AYP determination process. Although the Act calls for annual assessment in grades 3 through 8, the Act does not intend for the assessments to be used to track student progress longitudinally. As will be described in detail, on an annual basis the AYP determination process is solely interested in the percentage of students within a school (school district and state) who demonstrate proficiency. As a result, some schools will be identified as meeting AYP requirements and others as failing to meet AYP requirements.

Recently the Federal government did announce that it would consider longitudinal growth analyses as part of AYP but for only a limited number of states. Because of the maturity of Nevada's assessment system, it is contemplating the introduction of a growth component beginning in the 2006-07 school year. Until then, actual growth is not rewarded by the NCLB system of accountability. In fact, a school can show no change or actually show a decline in the percentage of proficient students and still be recognized as adequate while another school may demonstrate a significant positive change in the percentage of proficient students but still be recognized as less than adequate.

The AYP Determination Process

On an annual basis, schools, school districts, and the state as a whole must be judged regarding the adequacy of their progress. As discussed above, the inclusion of the word "progress" in the annual judgment process lacks precision. We are really talking about static performance and not progress. The basic question being asked by the model is if in the current school year a sufficient percentage of students are demonstrating satisfactory knowledge and skill attainment. If "yes," good; if "no," bad. As will be revealed, the

determination process is quite complex. There are a multitude of factors to consider before that final judgment can be made.

Subject Area Achievement Indicators

AYP is determined separately for English Language Arts (ELA) and Math. At this time, the federal mandate does not require the inclusion of science results. For each subject, the state must establish annual goals indicating the minimum percentage of students that must score at or above the "meets standard" level of achievement on the Nevada AYP tests. This is often referred to as the "percent at/above cut" or the PAC. The PAC is used to make status comparisons. If a school does not meet the PAC goal but demonstrates a decrease in the percentage of non-proficient students of at least 10% from the previous school year and meets the "other indicator" criteria, the school can be judged as meeting the AYP achievement indicator. This is referred to as the Safe Harbor provision.

Participation Indicators

Schools are required to have at least 95% of all students participate on the state AYP tests to meet the AYP requirements. Participation rates on ELA and math tests are considered separately.

"Other" Indicators

In addition to subject area proficiency and test participation, schools must be judged with respect to at least one "other" indicator. A school can fail AYP solely based on the other indicator and if it does this failure is tracked separately from ELA and math performance. Like the achievement comparisons, school performance on the other indicator is compared against an annual statewide goal. But the other indicator, as mentioned above, also comes into consideration if a school must use Safe Harbor to meet an achievement indicator.

At the high school level, the NCLB Act requires that graduation rate be used as an other indicator. The Act gives states flexibility in the use of other indicators at the elementary and middle school levels. State statute now requires that elementary and middle schools in Nevada be judged relative to average daily student attendance. Federal policy does not limit the number of other indicators that are used by a state, but it clearly states that other indicators can only be used to identify more schools as failing AYP and not to remove schools that would have otherwise failed with respect to the achievement indicators.

Subgroups

ELA participation, ELA achievement, math participation, math achievement, and, in certain instances (i.e. safe harbor analyses if necessary), other indicator performance are judged separately for up to 9 separate student groups. These include the school as a whole, five major race/ethnic subgroups (American Indian, Asian/Pacific Islander, Hispanic, African American, and white), students with disabilities, students with limited English proficiency, and students who are economically disadvantaged (economic disadvantage is predicated on eligibility for free and/or reduced lunch). It is important to note that individual students are counted multiple times when we consider subgroup analyses. For example, the majority of students in Nevada with limited English proficiency are classified as Hispanic. Many of

these same children are eligible for free and/or reduced price lunch, and they are counted as part of the whole school.

In brief, a school as a whole and each of its identifiable student groups has several hurdles to jump in order for the school to be deemed as demonstrating adequate yearly progress. First, a minimum of 95% of the students from each subgroup must participate on the state tests. Second, each subgroup must meet the achievement indicator (i.e. statewide status goal or the safe harbor provision) in each subject area. Third, the school as a whole must meet the other indicator criterion. The other indicator must also be passed for any student group that meets the achievement indicator through the Safe Harbor provision.

The state and school districts must make many comparisons for each school, and the school must pass each satisfactorily to be deemed adequate. By contrast, a failure with respect to a single comparison may lead to a negative AYP classification. Presented in Table 4 is a summary profile of the basic comparisons that must be made when determining AYP.

Table 4. Hypothetical School AYP Profile

Population	ELA Participation	ELA Achievement		Math Participation	Math Achievement		Other Indicator
		ELA Status	ELA Safe Harbor		Math Status	Math Safe Harbor	
School	Yes	Yes	NA	Yes	Yes	NA	Yes
American Indian/ Alaskan Native	Yes	Yes	NA	Yes	Yes	NA	NA
Asian/ Pacific Islander	Yes	Yes	NA	Yes	Yes	NA	NA
Hispanic	Yes	Yes	NA	Yes	Yes	NA	NA
African American	Yes	Yes	NA	Yes	Yes	NA	NA
White	Yes	Yes	NA	Yes	Yes	NA	NA
IEP	Yes	Yes	NA	Yes	Yes	NA	NA
LEP	Yes	Yes	NA	Yes	Yes	NA	NA
Low SES (FRL)	Yes	Yes	NA	Yes	Yes	NA	NA

As depicted in Table 4, 9 student groups were judged. The YES indicates that the hurdle was successfully jumped. Because the Status hurdle was successfully jumped by each group, the Safe Harbor hurdle was not applicable (NA). Because Safe Harbor was not applicable, the other indicator comparisons for groups other than the school as a whole were also not applicable. As can be seen from the table, there are 37 basic comparisons but this can grow to 63 comparisons very easily. In reality, many schools do not include all possible student groupings. Additionally, for most schools Safe Harbor is required to meet the achievement target and so the other indicator comparisons also become more relevant.

This general description gives a better sense of the complexity involved in judging a school; however, additional steps must be taken through analysis to properly judge schools. Next we discuss how each of the AYP variables is operationalized and how other key methodological considerations impinge upon the analysis.

Operational Definitions and Key Methodological Considerations

Full Academic Year

One of the initial steps in organizing information in preparation for the AYP analyses is to identify the population of students to be included. No Child Left Behind allows achievement indicators reflecting school and disaggregated group test performance to be based only on students who have been enrolled for a full academic year. By contrast, test participation and other indicator performance do not include this filter in defining the eligible student population. Instead, all students must be included. Hence, we need a definition of "full-academic year" and we need to be able to apply it to filter the included population at appropriate times.

- A student is considered to be enrolled at a particular school for a full academic year (FAY) if he or she has been continuously enrolled from the official count day of students, occurring in early fall, until and during the specified test window, which occurs in mid-Spring.[3]

Data Aggregation

A second key issue is the requirement that states build reliable and valid systems of determinations. Much debate has ensued regarding how or what approaches are best suited to support the reliability of the system. There are two assumptions that predicate some of the choices made in Nevada. First, the more individual comparisons made to profile a school, the greater probability that a school will be identified as failing AYP. Second, the greater the proportion of students within a school included in the assessment system, the greater the probability that the final AYP determination is representative of the school.

Following this logic, and to the benefit of schools, data is aggregated across grades when making AYP determinations. For example, the most typical grade configuration among elementary schools in Nevada is a K-5 structure. As noted above, assessments included in the 2005-06 determinations are administered in grades 3 and 5. Therefore, all assessment-related indicators reflect an aggregation (adding) of students across grades 3 and 5.

Looking back at the assumptions, aggregating across grades means far fewer school level comparisons (in contrast to grade by grade comparisons) and a more representative population with some control over the effect one particular cohort of students may have on

[3] Note that the same rules apply to school districts. A student is considered in the achievement-based analysis for school districts if he or she has been in the school district for a full-academic year. The difference is that a student could be included in a school district analysis even if he/she had attended two or more schools during the school year, so long as the different schools are all part of the same school district. For the state AYP analysis, all students, regardless of years in school/years in district are included in the analysis.

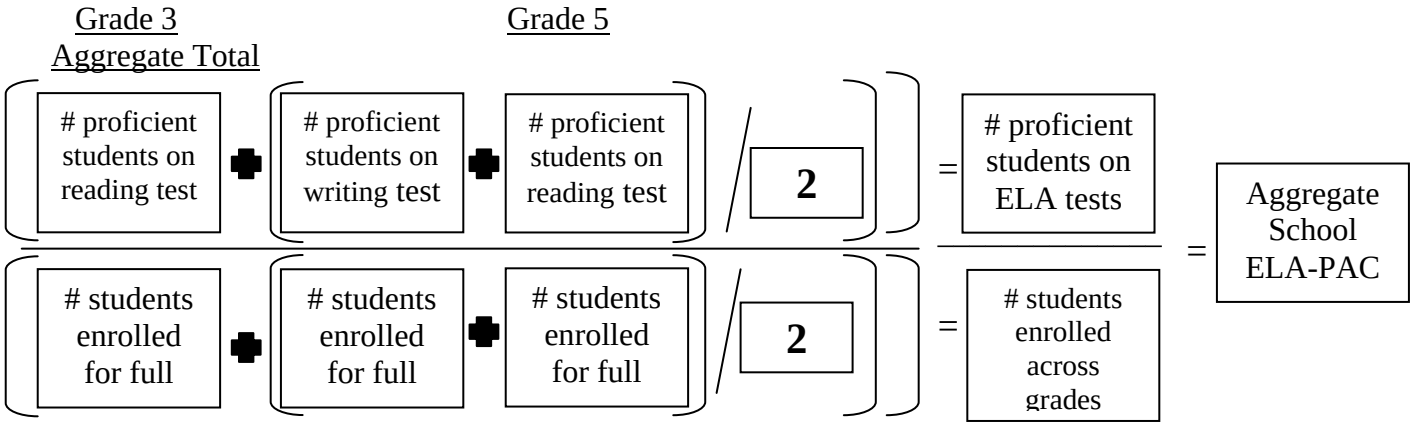
the whole school (i.e. a poor performing 3rd grade cohort may be offset by a higher performing 5th grade cohort).

Data aggregation is not accomplished easily. It involves more than simply adding students but requires consideration of combinations of different tests. For example, for the 2005-06 determinations, the elementary aggregation includes math tests in grades 3 and 5, but for ELA we have reading tests in grades 3 and 5 and a writing test in grade 5. Additionally, some IEP students will take the NASAA alternative, and it is possible that school districts may use a content alternative for LEP students.

This means that aggregation will include 2 grades and several different tests. Combining information among the disparate tests is complicated by different ways those tests are scored. For example, the reading test is scored on a scale ranging from 100 to 500 while the 5th grade writing test is scored on a scale ranging from 4 to 20. These scale differences are overcome by combining the tests using the previously described achievement level scores. As noted above, each test used in the system must align to state content and achievement standards. The tests align to the achievement standards by yielding achievement level scores (i.e. meets standard). These scores provide a general statement regarding a student's overall performance relative to Nevada standards.

Although the reading test at grade 5 taps different content standards than the 5th grade writing test, each provides an estimate of student proficiency relative to ELA standards. We could simply add up the number of "meets standard" students and divide this number by the total number of test takers, but this would in effect double count 5th grade students. To control for this, we divide this group in two before adding 3rd grade students to the mix.

Generally speaking, we combine the numbers of proficient students for each test and divide that number by the aggregated grade level enrollment counts. Graphically the aggregation looks as follows:



In short, data aggregation is one of the measures employed to meet the NCLB requirement of a reliable and valid system. Similar steps must be taken in middle school and in high school to derive PAC estimates. Through aggregation we minimize the number of comparisons a

school is subjected to, provide a better proportional representation of the school, and, hence, increase the stability or reliability of the data used to make comparisons.

A third consideration, also very much related to system reliability and validity, deals with the identification of student groups and the confidence associated with student group test performance. We discuss these issues next.

N-Size, Confidence Intervals, and Systems Reliability & Validity

As noted above, it is assumed that greater proportional representation of a school enhances the reliability of school performance estimates. This should not be taken to imply that some specific number of students is required to reliably represent a school or subgroup within a school. NCLB requires that states employ reliable and valid systems and makes several references to minimum group size for reporting and comparison purposes. Unfortunately, conventional statistical wisdom and practicalities associated with public schools forbid a simple application of a minimum n-size if both reliable and valid interpretations are being sought.

Much national debate on this issue has ensued. Within the debate, it is suggested that a minimum n-size to "guarantee" some reasonable degree of reliability might be as low as 100 students or as high as 350 students. Clearly this sort of a rule would result in the exemption of large numbers of schools from standard statistical comparisons of their performance and would eliminate consideration of almost all subgroup comparisons in almost all schools. Although the large n-size might provide some confidence in the achievement estimates, the exclusion of large numbers of schools and entire student groups within schools severely threatens the validity of the system. This was not the intent of the law ("leave no child behind").

There is legitimate reason to be concerned when sample sizes become very small. With very small samples, estimates from year-to-year are very unstable, and observed shifts in performance, unrelated to school effectiveness, are likely to occur. So, a desirable balance is to be sought between too many and too few students.

One solution lies not in the establishment of a specific "n-size" but in the employment of statistical controls. The use of confidence intervals is one such approach. Confidence intervals can be used productively to rule against year-to-year instability created by factors extraneous to school effectiveness (e.g. differences between cohorts of students unrelated to instruction). One benefit of this approach is that minimum n-size requirements can be eliminated or set very low. Confidence intervals can be computed with sample sizes as small as 5.

At first glance, it seems that confidence intervals are useful in order to deal with small samples of students. But in reality, the decision to use them is less based on sample size and more based on desire for "error-free" estimates. Error always affects achievement estimates, but confidence intervals allow control for some of the known contributors to measurement error (i.e. sampling error). The degree of confidence is predetermined prior to conducting the

statistical tests. In this way, the same degree of confidence can be achieved when our n-size includes 100 students in a school or when it includes 25 students. This also enables us to have the same confidence for multiple student groups within a school even though their sizes may vary.

There has been significant outcry regarding the minimum n-size issue and many seem reluctant to rely solely on statistical approaches to control "error." Because of this, Nevada has chosen to use a hybrid approach in which it employs confidence intervals but does not conduct statistical tests on subgroups within schools with fewer than 25 students in the aggregate (e.g. summed across grades 3 & 5).

How do confidence intervals work?

Confidence intervals and their application are built on the basic measurement assumption that all measurements contain random error. In other words, an observed performance on a test is equal to "true" performance plus the effect of random "error." Examples of random error might include a dog barking outside the window of the testing room, having no air conditioning, being sick, using a test form that was positively biased in terms of the content most recently studied, or making a lucky guess.

This means that for any given test administration, a student's observed score is as equally likely to be an overestimation as an underestimation of the student's "true" level of achievement. Sometimes our test scores suggest we are more knowledgeable than we actually are and at other times test scores suggest we are less knowledgeable than we actually are. Confidence intervals allow us to specify the "limits" within which true performance may fall. They allow us to set the upper and lower limits of performance estimation. For example, if a student scored a 50 on a test, confidence intervals allow us to know the likelihood that the student's true score is between 40 and 60.

The breadth or width of the "limits" depends on how much confidence we desire. Greater confidence results in a greater interval width. For example, if we score 50 and we are not too concerned about the accuracy of our judgment, we may expect that the true score lies between 45 and 55. If we are more concerned and desire more confidence, we may use 40 and 60. In other words, there is less chance that we would be making an error if we use the 40 to 60 range than the 45 to 55 range.

Using a graph of a normal distribution (e.g. bell-shaped curve), it is easy to see the relationship (see Figure 1). We can look at the normal distribution of scores around the observed score and specify the amount of confidence being sought. In this example, a two-tailed 95% limit is illustrated.

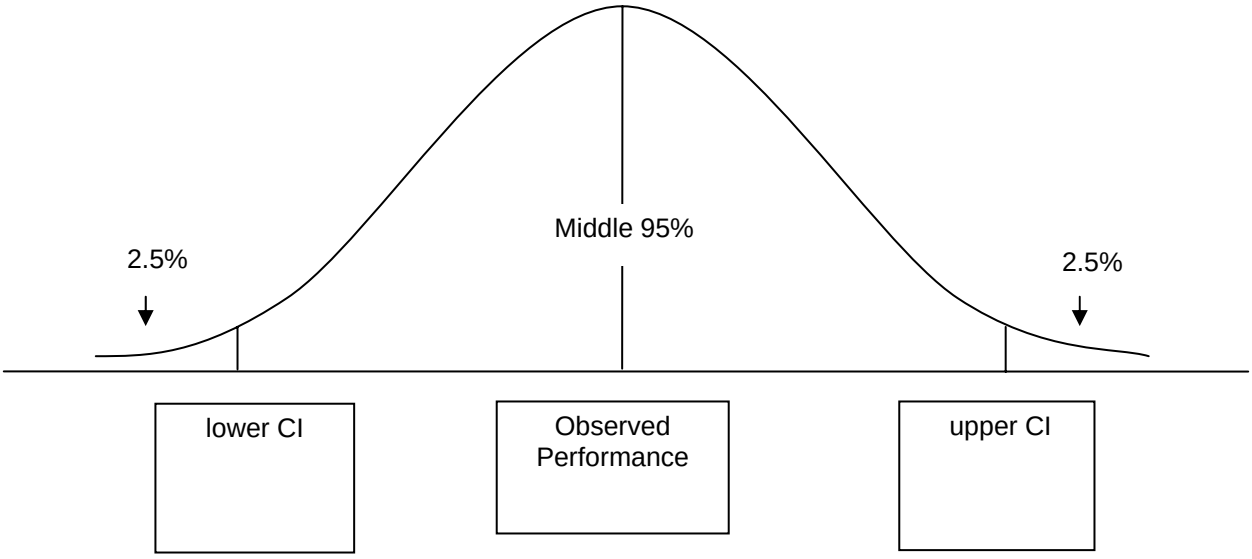


Figure 1. A Normal Distribution

The graph depicts how the observed score may underestimate performance or overestimate performance. By organizing decisions using a two-by-two classification table (Table 5), we can see the sorts of errors that can be made by using observed performance without considering measurement error (e.g. false negative and false positive classifications).

Table 5. Potential Classification Errors

	High Test Score	**Low Test Score**
High Achiever	Correct	*False Negative*
Low Achiever	*False Positive*	Correct

Table 5 depicts the classification of high and low achievers based on test performance. From it we see that some high achievers score high on the test and others score low. When they score high we correctly classify them, and when they score low we incorrectly classify them. This error is referred to as a "false negative." By contrast, a low achiever may score high or low. A low achiever who scores high will be incorrectly classified as high achieving. This is referred to as a "false positive."

There is always some likelihood that both errors can occur but attempts can be made to control for them. When emphasis is placed on one of the error types (i.e. false negatives) there is greater likelihood that the other error will be committed. Which error type receives greater control is a matter of choice and is often predicated on the consequences of the decision. For example, if we give accolades to high achievers and reward their achievement, we might want to be careful to be sure that the accolades are deserved, and so we control

more against false positives. By contrast, if we sanction or punish low performance, we may want to be careful that the sanction is deserved, and so we control more against false negatives. Given the consequences associated with NCLB, the choice has been made to emphasize control against false negatives. In practice this means that we focus our attention on the upper limit of observed performance when we make comparisons.

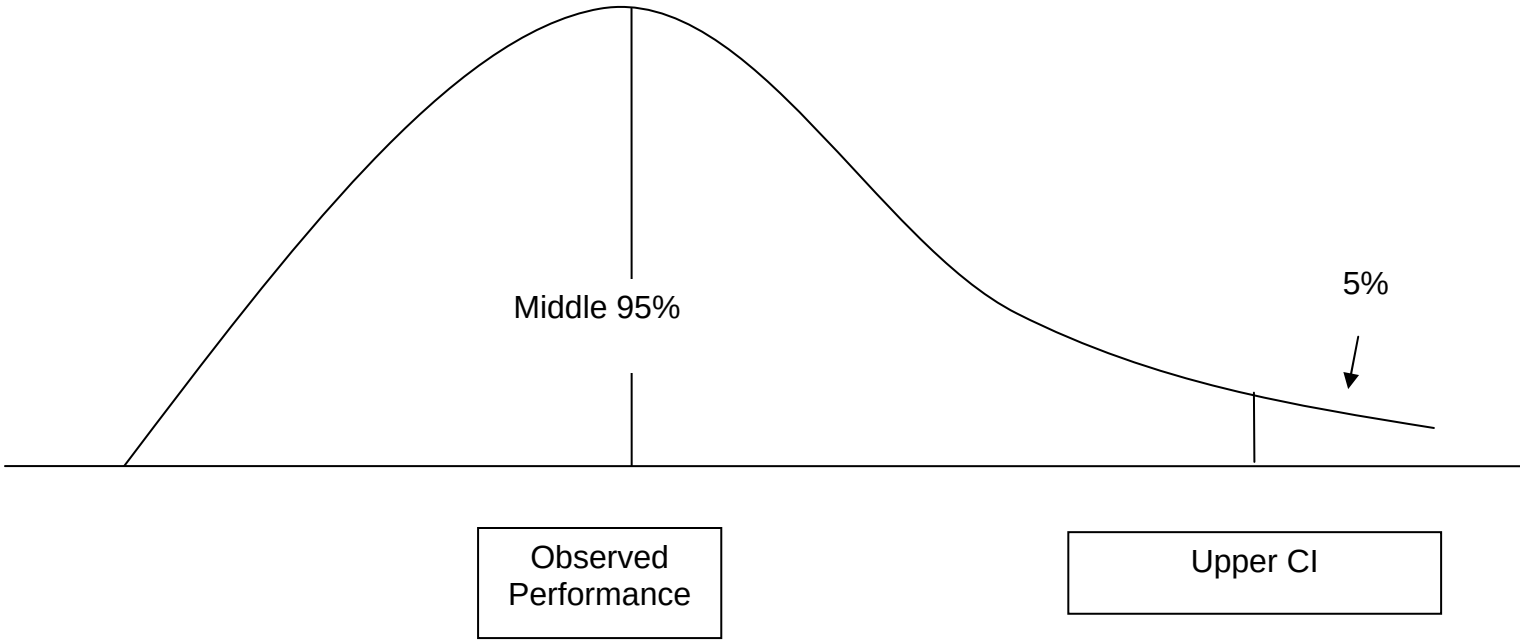


Figure 2. Negatively Skewed Distribution With Focus on Upper Boundary

By doing this we have made the choice to accept more false positive errors. This is a tough choice but prudent given the consequences associated with negative AYP classifications. In other words, before applying invasive corrective actions to a school, we want to have a high degree of confidence that we have not falsely identified the school as failing. It is important to restate that this application of control can be done with virtually any sample size, establishing an "even" playing field.

It is understood that the use of statistical methodology can be a challenge. However, unless we accept the infallibility of "measurement," it is obligatory to account for unreliability when we make important decisions that ultimately impact the lives of students and educators. A conceptual understanding of measurement error prompted by graphic illustrations also underscores how a narrow focus on sample size misses the point. Even with large samples, errors can be made. The point is that regardless of sample size, the use of confidence intervals better allows us to achieve reliable or consistent decision making.

Operational Definitions : Achievement, Participation, & Other Indicators

We now address each AYP indicator in turn, beginning with the achievement indicators and ending with the other indicators, we demonstrate the basic calculation for each and explore how each is used in the AYP determination process.

Achievement Indicators

Achievement Indicator 1: Status / PAC

The PAC (Percent at/above Cut) for a school is the primary achievement indicator for determining AYP. It is derived by simply dividing the number of students who demonstrate proficiency or performance indicative of meeting or exceeding standard by the total number of students.

$$\frac{\text{Number of meets or exceeds standard students}}{\text{All students}} \times 100$$

For example, if 1000 students were in the enrolled population for an elementary school, and 389 scored at or above the meets standard cut score, the PAC would equal 389/1000 x 100 or 38.9%. Note that in making comparisons, we round percentages to the second decimal place. So if the total population had been 900, PAC would equal 389/900 x 100 or 43.222222222222… (or 43.22%).

Achievement Indicator 2: Safe Harbor / Relative Growth

Safe Harbor, or relative growth, refers to the percent reduction in the percentage of non-proficient students. It relies on the calculation of PAC but requires the PAC in the current school year to be compared to the PAC from the previous school year (e.g. percent differences in PAC rates). So, using the same definition above for calculating PAC, relative growth is equal to:

$$1 - [\ (100 - PAC_{\text{current year}}) / (100 - PAC_{\text{previous year}})\ ]$$

For example, let's assume that in the current year the PAC was 30.00% and in the previous year it was 25.00%. Relative growth would be equal to 1 – [(100 – 30.00) / (100 – 25.00)] or 1 – (70.00 / 75.00) or 1 - .9333 or .0667. This means that the percentage of non proficient students was reduced by 6.67%.

Participation Indicator

Participation rate is an easy indicator to derive. It is equal to the number of students who took the test divided by the total number of enrolled students.

$$\frac{\text{Number of students who took the test}}{\text{All students}} \times 100 \text{ (rounded to second decimal place)}$$

For example, if 950 students took the test and 1000 students were enrolled, participation rate would be equal to 950/1000 x 100 or 95%.

The difficulty in determining participation rate is not in calculating the rate but in determining what constitutes "participation." At this juncture, we consider any student who has made a legitimate attempt at taking a test or one of several subtests as having participated. The only way for us to determine this is by observing at least a single valid response to a test.

Other Indicators

Other Indicator 1: Average Daily Attendance (ADA)

Average Daily Attendance is also an intuitive and relatively easy indicator to calculate, assuming the necessary data elements are available. Considering and collapsing across all students in the school or subgroup within the school, we divide the sum of days present by the sum of days present plus days absent:

$$\frac{\text{Sum of days present during the school year}}{\text{(sum of days presents) + (sum of days absent)}} \times 100 \text{ (rounded to second decimal)}$$

To calculate this figure, we must know for each student the number of days they were considered "in attendance" and the number of days they were absent. Because of the timing of when this calculation must be made and because of significant differences in school schedules (e.g. multi-track schools), we estimate average daily attendance based on the first 100 days of instruction within the school. Data to complete these comparisons are furnished by school districts and are based on attendance information pulled directly from their local student information systems.

Other Indicator 2: Graduation Rate

Graduation rate is calculated to represent the school as a whole or any of its subgroups. Graduation rate is quite complex in its calculation. Its calculation is dependent on the availability of several pieces of information.

Graduation rate involves the percentage estimate of graduation among a cohort of students (e.g. the graduating class of 2005). Although longer time periods can be considered, in its initial calculation graduation rate assumes the traditional 4-year high school schedule. Information must be collected on a cohort of students throughout those four years to generate

the rate. This includes annual dropout rate figures, and figures pertaining to several separate completion options (e.g. standard diplomas, adjusted diplomas, certificates of attendance, GED recipients). Once collected, the basic formula involves the division of students receiving standard diplomas by a combination of all completion possibilities, along with the numbers of students dropping out of school in grades 9, 10, 11, and 12. The formula for calculating graduation rate is as follows:

$$\frac{\text{Number of standard diploma recipients}}{\text{Number of standard diploma recipients} + \text{Number of adjusted diploma recipients} + \text{Number of certificate of attendance recipients} + \text{Number of GED recipients} + \text{Number of } 9^{th} \text{ grade dropouts} + \text{Number of } 10^{th} \text{ grade dropouts} + \text{Number of } 11^{th} \text{ grade dropouts} + \text{Number of } 12^{th} \text{ grade dropouts}} \times 100 \text{ (rounded to second decimal place)}$$

As demonstrated from the formula, much information about a cohort of students is needed to calculate graduation rate. Most of the information is currently available for the calculation of graduation rate for whole schools and for race/ethnicity subgroups. At this juncture, sufficient historical data has not been collected on dropouts or on some forms of completion for students with disabilities, students with limited English proficiency, or for economically disadvantaged students. Organized collections with respect to GED recipients are now under way as a consequence of the federal requirements. It is anticipated that the 4-year cohort information for these student groups will be available for introduction during the 2006-07 school year. When graduation rate is unavailable for a student group, we must use average daily attendance instead.

Finally, two substantive differences exist between graduation rate and the other AYP indicators. First, the need for completion information makes it impossible to calculate the indicator for the "current" school year. In other words, for the 2005-06 AYP determinations, graduation rates reflecting the graduating class of 2004-05 have to be used. Second, the graduation rate indicator collapses information across a four-year time span, while the other indicators rely primarily on a single year of information. This means that change with respect to graduation rate is likely to take more time to observe. In other words, the introduction of an intervention by a school to address dropout issues is likely to take a considerable amount of time to affect the actual graduation rate for a school or any of its student groups.

Indicator Comparisons: Achievement, Participation, & Other Indicators

Now that we know how the AYP variables are derived, we now need to look at how they are used in making the comparisons illustrated in Table 4. We look at each indicator in turn.

Achievement Indicators

Status/PAC Comparisons:

As part of the AYP determination for a school, the PAC rate for the whole school and for each of its identifiable subgroups must be compared against an associated statewide annual measurable objective (AMO) PAC target rate. As noted above, the PAC comparisons are made separately in ELA and math using only students enrolled within the school for a full academic year or longer.

Although different PAC targets exist in ELA and in math, and therefore comparisons must be made separately in each subject area, the same PAC goal must be used to judge whole school and student group performance. Moreover, subject area PAC goals are different for different clusters of schools. In Nevada, with few exceptions, schools are clustered in three ways: elementary schools, middle schools, and high schools.

In requiring the establishment of PAC AMOs, the federal government specified that the initial targets be based on baseline estimates of PAC established from assessments administered during the 2001-02 school year. Additionally, in establishing targets for the 2002-03 school year and beyond, states had to build targets to ensure that 100% of students would be proficient by the end of the 2013-14 school year. States were given the option to hold rates constant over predetermined time periods. However, changes in annual targets leading to 100% proficiency had to be equidistant.

States were provided two methods to establish baseline annual measurable objectives. However, states were required to use the method that yielded the higher initial PAC rate. This resulted in the use of what has been termed the “school percentile method.”

Table 6. School Enrollment Percentile Method

School Name	Percent Proficient (PAC)	Enrollment	Cumulative Enrollment Percentage
School A	**2%**	**127**	**0.4%**
School B	**3%**	**28**	**0.5%**
School C	**6%**	**56**	**0.7%**
School D	**7%**	**380**	**1.97%**
School E	**10%**	**150**	**2.5%**
School X	**40%**	**281**	**20%**
State Total		**30000**	**100%**

The school percentile method involves the ranking of schools in terms of subject area PAC rates from lowest to highest. In addition, the student enrollment for each school is indicated. The PAC rate for the school at the 20th percentile of cumulative state enrollment is chosen as the baseline rate for the given subject. Table 6 illustrates this method. Based on it, the PAC rate associated with school X would be used for baseline. Using this method, Table 7 includes the PAC AMOs for the 2005-06 school year.

Table 7. PAC AMOs for 2005-06

	ELA	MATH
Elementary	39.6%	45.4%
Middle School	47.5%	43.3%
High School	77.9%	52.3%

When submitting its AYP plan to the federal government, the state was required to describe its method for establishing baseline and its estimated targets for each school year leading up to the 2013-14 school year. The state has used a tiered approach in which it holds constant annual goals for as long as allowed by the federal government, meanwhile making equidistant target increases when necessary. Figure 3 depicts the state's tiered approach, and Table 8 includes the estimated AMOs through 2013-14.

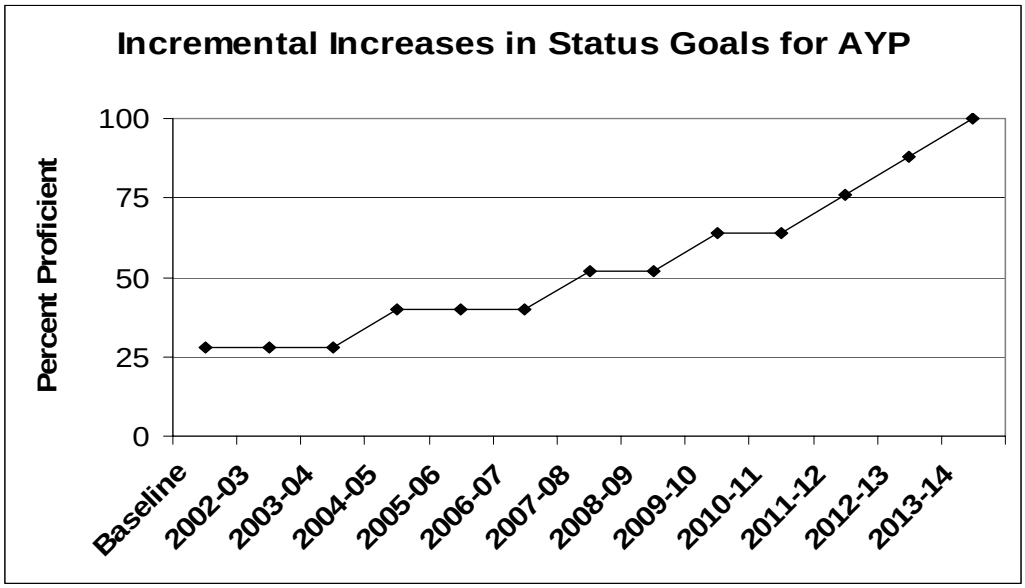


Figure 3. Tiered Approach to Adjusting State PAC AMOs

The annual goals depicted in the graph are for elementary ELA, but the scheduled tiered increases reflect Nevada's plan for each subject area at each grade span level. For the 2002-03 and 2003-04 comparisons, annual targets were established by holding baseline estimates constant. The first increase required by the federal government occurred for the 2004-05 school year. That goal will be held constant for two additional years followed by an increase for the 2007-08 school year. Given the equidistant requirement, the target can only be held constant for one additional year before an increase in 2009-10. That target is again held constant for one additional year followed by increases in each of the last three years of the NCLB 12-year accountability time span. As depicted in Figure 3, there are six increases in the 12 year period. For any subject at any grade cluster, the 12-year plan can be established by subtracting the baseline PAC rate by the 100% goal and dividing that difference by six. The resulting dividend is the necessary increase at each of the six time intervals.

Table 8. Estimated Annual Measurable Objectives Through 2013-14 School Year

School year	Elementary School		Middle School		High School	
	ELA	Math	ELA	Math	ELA	Math
Baseline, 2002-03, 2003-04	27.5%	34.5%	37%	32%	73.5	42.8
2004-05, 2005-06, 2006-07	39.6%	45.4%	47.5%	43.3%	77.9%	52.3%
2007-08, 2008-09	51.7%	56.3%	58%	54.6%	82.3%	61.8%
2009-10, 2010-11	63.8%	67.2%	68.5%	65.9%	86.7%	71.3%
2011-12	75.9%	78.1%	79%	77.2%	91.1%	80.8%
2012-13	88%	89%	89.5%	88.5%	95.5%	90.3%
2013-14	100%	100%	100%	100%	100%	100%

The title of Table 8 indicates that annual objectives are estimated. As noted, the federal government required that states immediately implement the new accountability systems

relying on those assessments available during the 2001-02 school year. Furthermore, states were required to implement annual assessments in grades 3 through 8 and at the high school level by the 2005-06 school year. The implication of this in Nevada is that the assessment system used to establish baseline in the 2001-02 school year will undergo significant change until full implementation occurs in 2005-06. It is reasonable to expect some revision after that as well.

To address the changing assessment landscape, the Nevada plan submitted to the federal government indicated that as significant revisions occurred to the state assessment system, revisions to annual AYP targets could result. In no case can changes result in anything other than the 100% expectation in 2013-14, but the annual targets leading to that 12-year goal could change.

The first known change to the annual targets occurred in the 2003-04 school year. During 2003-04, criterion-referenced tests were implemented at grade 8 and resulted in the removal of a 7th grade norm-referenced test that had been used in determining AYP. Similarly, a 4th grade NRT has been removed from AYP consideration and was replaced by the 4th grade writing test. Because baselines were so heavily dependent on the inclusion of NRTs, it has been judged prudent to reestablish school targets if the data supports a change. Based on analyses undertaken following the receipt of the 2003-04 data files, slight adjustments were made to the AMOs at the elementary level, but no adjustments were required at the middle school level. A second major revision affecting both elementary and middle school targets is expected to occur in 2005-06 or 2006-07 as the full implementation of grade 3 through 8 tests occurs. The writing test previously administered at grade 4 has been moved to grade 5. No changes to AMOs will result from this assessment change. However, 2005-06 is the first year that reading and math tests are administered "live" in grades 4, 6, and 7. Because of the timing of standard setting (e.g. establishment of test cut scores) for these tests, no change to the 2005-06 AMO is anticipated. However, changes to AMOs beginning in 2006-07 may occur.

The significantly higher PAC AMOs at the high school level are worth mentioning. The basic difference between high school and the other school levels is in how the assessment results are used in the AYP calculations.

The federal regulations supporting the NCLB assessment provisions prescribe that for AYP purposes, states only use students' first testing opportunity. In other words, if a state administers a 3rd grade test more than once in grade 3, it is required to use the first administration of that test in determining AYP.

This requirement created a difficulty for states such as Nevada that use high school exit examinations and have chosen to include those in the AYP system as opposed to creating a separate set of accountability measures. The "exit" examinations in Nevada are used to measure proficiency as students are exiting high school. However, in the service of fairness, the state begins providing opportunities for students to take these examinations as early as grade 10. It would be inappropriate to rely solely on 10th grade performance when we know

that the test is designed to measure content that students might not be exposed to until after that time period.

Because of this situation, the Nevada plan requested the use of a cumulative exit examination pass rate in establishing annual PACs. The request proposed including up to five test opportunities through April of a student's 12th grade year of instruction. That request was denied, but the federal government allowed the inclusion of the first two student attempts. Still, the allowance of two attempts results in higher PAC rates in high school.

Returning to a discussion of the status PAC comparisons, each year the PAC for the whole school and for each student group is compared against the associated statewide AMO for the school level and subject area. For example, in elementary school the ELA PAC annual measurable objective in 2005-06 is 39.6%. To meet the status requirement, the whole school and each subgroup has to have an ELA PAC rate of at least 39.6%.

To strengthen the reliability of the status comparisons, the standard error of the proportion is used to adjust observed PAC scores before a comparison against the state target is made. This is an important step allowing for a predetermined degree of confidence in the status comparisons (See discussion of confidence intervals above).

The standard error of the proportion accounts for sampling error associated with a particular cohort of students. The formula for the adjustment is relatively simple to apply. We multiply the proportion of proficient students by the proportion of non-proficient students and divide that quotient by the number of students in the sample. We then take the square root of that dividend.

$$\sigma_P = \sqrt{PQ / N}$$

For example, if 120 students in the school took the ELA tests and 36 scored at or above proficient (30% PAC), we would multiply 30% by 70% and divide that figure by 120. This equals .0027. The square root of .0027 is .0459. The result is a single standard error. The single error adjustment provides more certainty than relying on the observed score alone, but the level of certainty might be considered low using conventional applications of statistical tests. To ensure at least a 95% level of confidence, we can multiply the standard error by 1.645 (e.g. one-tailed z-score transformation). Applying this rule to our standard error results in a one-tailed adjustment of .0459 x 1.645 or .0754 (7.54%).

We add the 7.54% adjustment to the observed PAC of 30.00%. Thirty plus 7.54 is equal to 37.54%. The adjusted PAC of 37.54% is the figure compared against the state annual measurable objective (AMO) of 39.6%; in this case the school would be deemed to have not met the status achievement comparison.

The use of the adjusted score is not intended to give schools an advantage or to lower expectations. As stated, the use of the confidence interval is intended to better ensure the reliability of the system by warding against false negative classifications. Having a reliable

system is a federal requirement. By making the adjustment as noted above, we have a relatively high degree of confidence that the school's "true" level of performance is below the state target and, hence, the application of sanctions is more defensible.

Safe Harbor/Relative Growth Comparisons:

When schools or subgroups fail to meet the PAC AMO, they can still be deemed as making AYP if the percentage of non-proficient students is reduced by 10% or more and if the group in question meets the criteria of the other indicator analysis. Therefore, the Safe Harbor/ relative growth comparison is a conjunctive analysis in which **both** a 10% reduction in the percentage of non-proficient students **and a** performance exceeding the target on the other indicator is observed (see the next section for information on other indicator comparisons).

As with status/PAC comparisons, only students enrolled for a full-academic year are included in this comparison.

For this comparison, the NON-PAC (e.g. percentage of non-proficient students) rate from the current year is divided by the previous year's NON-PAC rate. The resulting ratio is then subtracted from 1 to obtain the observed percent reduction in the percentage of non-proficient students last year. The change is compared to the 10% change threshold. As with status comparisons, the state uses confidence intervals to ensure more reliable decisions. Because the PAC rate from two separate administrations is considered, control over sampling error from both administrations must be achieved, and so a slightly different error estimation formula is used.

With this end in mind, the standard error of the difference in proportions is used. The formula is similar to the formula for the standard error of the proportion. The basic difference is that we take the square root of the sum of the variance estimate for each separate administration.

$$\sigma_{P1-P2} = \frac{\sqrt{\sigma_{P1}^2 + \sigma^2_{\text{P2}}}}{\text{NON-PAC}_{\text{Previous Year}}}$$

To help explain the Safe Harbor calculation, we provide an example. In the current year 36 of 120 students or 30.00% were proficient (PAC) whereas in the previous year 25 of 100 or 25.00% of the students were proficient. We calculate the change in the NON-PAC rate by dividing the current year NON-PAC (e.g. NON-PAC = 100-PAC) rate by the previous year NON-PAC rate and then subtracting the ratio from one. This works out as:

1 - [(100 - 30.00) / (100 - 25.00)] = 1 – (70.00 / 75.00) = 1 - .9333 = .0667.

The result suggests a 6.67% reduction. This difference still must be adjusted to account for sampling error.

The adjustment is derived using five steps. First, multiply the proportion of proficient students by the proportion of non-proficient students and divide that result by the number of participating students to obtain the variance estimates. This must be done for both years in question. Second, sum the variance estimates. Third, take the square root of the summed variance estimates. Step three yields the standard error of the difference in proportion. Fourth, multiply the standard error by the appropriate z-value to establish the confidence limit. The federal government required that Nevada cap the confidence associated with Safe Harbor comparisons at .75. A z-score of .675 establishes this limit. Finally, divide the resulting confidence interval by the percentage of non-proficient students in year one to convert the confidence interval for the difference in proportions to the confidence interval for the percent difference in proportions.

So for our working example:

Standard error = Square root of { [(.30 *.70) / 120)] + [(.25 * .75)/100)] } = .0602
Z-score transformation to .75 limit = .0602 * .675 = .0410.
Conversion to CI for percent difference in proportions = .0410 / .75 = .0547

To adjust our observed difference we add the confidence interval to the observed difference or .0667 + .0547. This equals .1214 or 12.14%. This number would be compared to the 10% change Safe Harbor threshold, and in this case we would judge the school to have met the Safe Harbor criterion.

Test Participation Rate Comparison:

As a reminder, all students enrolled at the time of testing must be included in the participation rate calculation. Participation rate must be calculated separately for ELA and math. No correction for measurement error is applied to this comparison.

The observed participation rate is compared against the 95% participation rate threshold (established by the NCLB legislation) for the whole school and for each of its student groups. This is a simple comparison, but 95% is a rigorous standard. Moreover, if strictly applied for a school or subgroup within a school with 20 students, all but one student would have to participate to meet the criterion. If the school or subgroup had 19 or fewer students, all students would have to participate to meet the 95% criterion (e.g. 18/19 = 94.7%).

There are legitimate circumstances that can result in a student's failure to participate. In an attempt to be sensitive to this, a modified criterion has been established for schools or subgroups within schools that have fewer than 20 students. It is labeled the N-1 rule. Instead of using the 95% threshold in this instance, we apply a standard of N-1, with N being equal to the number of enrolled students at the time of testing. For instance, if a school had only 19 students, at least 18 of the students would have to participate. Likewise, for a school with 10 students, at least 9 would have to participate.

Other Indicators

Other Indicator Comparisons:

As a reminder, all students enrolled at the school during the school year are included in this calculation. Additionally, no measurement error correction is currently used with these comparisons. As noted, only the school as a whole, not ethnic or special student groups (i.e. IEP, LEP, FRL), is judged against the other indicator target as a stand alone AYP analysis. Student group performance on the other indicator is only considered if a Safe Harbor comparison is needed when the PAC AMO was not met by the student group in question.

The first step was to establish state goals for these indicators. For average daily attendance (ADA), Nevada statute requires a 90% student attendance rate. The state adopted this threshold as its statewide objective to be reached by the school as a whole and by each identifiable student group when applicable.

For graduation rate, no such standard was established by statute. To explore alternatives, the Department of Education applied methods provided by NCLB to establish PAC indicator baselines. This resulted in an application of the school percentile method described previously. Using this as reference, the State Board of Education has temporarily adopted a graduation requirement of 50%.

For either ADA or graduation rate, a school or student group can meet the requirement in two ways. First, if they perform at or above the threshold, they have met the AYP requirement. Second, if they are below the threshold but have made some positive gain in comparison to the previous year, they are deemed to have met the criterion.

For example, if an elementary school's ADA is at or above 90%, they have met the other indicator criterion. If it is below 90% but it is greater than its ADA in the previous school year, it has met the other indicator criterion.

Section Summary

In this section, the key variables to be considered in making AYP determinations were outlined. Operational definitions of key variables, the establishment of statewide indicator goals and measures used to increase the reliability of our decisions were discussed.

Putting all of this together allows the state to conduct school and district level comparisons and to make preliminary AYP determinations. The next section outlines the transition from preliminary determinations to final determinations.

School and School District AYP Classifications and Annual Achievement Designations

Schools and school districts are judged annually and are classified as having met or having failed to meet AYP. Based on AYP classifications, schools and school districts receive

achievement designations. Designations include an identification of schools that are in need of improvement, high achieving, exemplary in achievement, and that are most improved. Schools that receive none of these distinctions are considered to be adequate in regard to achievement.

NCLB gives the ultimate authority for making school level AYP classifications to the school district. It gives the state the authority to make school designations based on the AYP classifications. NCLB gives the state the authority to make both school district AYP classifications and achievement designations. Nevada's Senate Bill 1, passed in 2003, mimicked the NCLB language providing similar authority. However, it requires the state to make all preliminary AYP classifications and requires school districts to consult with the Department of Education before making final AYP classifications.

School AYP Classifications

Based on data collected from test score sheets and limited information provided from other data sources (i.e. other indicator performance), the state conducts AYP analyses for all schools and school districts. The result is a profile for each school that summarizes the states preliminary findings.

The state issues the preliminary information to local school districts for distribution to their schools. The information is considered preliminary until both schools and school districts have ample opportunity to review the analyses and determinations.

Although legally schools are given an opportunity to appeal achievement designations, they are required to appeal their AYP classifications to the school district. As noted, school districts consult with the Department in judging appeals and prior to making final determinations. To this end, school districts must furnish the state with comprehensive support materials to assist in consulting on school level appeals (i.e. student data files and output from reanalysis of data). Subsequent to the consultation, school districts make the final classifications and must provide to the state a final listing of schools classified as not having met AYP.

The opportunity for appeal is given in the interest of making valid decisions. For example, a school may believe that the state analysis is in error for a variety of reasons, or it may believe that other evidence could be introduced that paints a different picture of school performance.

The NDE and local school districts have worked jointly to specify grounds for appeals. This is reviewed annually and shared with school districts prior to the receipt of preliminary findings. Although it is the ultimate authority of each school district to classify its own schools, there is interest in seeing an equitable and common application of the process across school districts. School districts retain the authority to classify a school even if the State Department disagrees with an appeal. In short, appeals may be granted if student performance was adversely affected by extraordinary and unavoidable circumstances during

testing, if significant coding errors impact the AYP analysis[4], if additional statistical analyses conducted by the school or school district identify errors in the state calculations, or if other significant factors produce statistical or substantive explanations for school performance.

School District Classifications→ At the same time that the state issues school level preliminary classifications, it must issue preliminary school district classifications. The school districts then have an opportunity to appeal their classifications directly to the NDE.

The NDE employs a nearly identical list of appeal grounds in judging school district appeals. When considering school district appeals, the NDE requires that school districts provide backup materials supporting their appeals.

After the close of the district appeal window, the NDE releases its final school district classifications by category. Table 9 includes an outline of the general timeline from the spring test windows to final AYP determinations to implementation of school/district improvement plans.

Table 9. Approximate AYP Schedule of Activities

April 15th	Approximate close of spring testing window
April 22nd	Score materials to test vendor
May 20th	Assessment reports to school districts
June 15th	NDE issues preliminary school and school district AYP classifications
July 1st to July 31st	Designation appeal window
August 1st	Final designations made

Achievement Designations

In Need of Improvement (INOI) Designations

Once AYP classifications have been formally determined, the tracking of schools ensues. As noted earlier, schools must be judged separately in ELA and Math as a requirement of NCLB. Following this logic, Nevada tracks schools by subject areas and by the other indicators separately. This becomes significant as we consider achievement designations.

To be designated as INOI a school must fail AYP in two consecutive years. In Nevada, this means that the failure must occur in the same subject area or relative to the other indicator in consecutive years. To assist in the designation process, schools are placed on what has been termed "watch" lists. The lists are distinguished by which subject area criterion, or by the other indicator criterion, schools failed to meet.

[4] Coding errors may occur for a variety of reasons. Those judged to be legitimate errors, and not errors due to negligence or errors that are repetitive from past years, may result in a reanalysis of AYP. Reanalysis is the responsibility of the school and school district.

Table 10. Single Year Classification Table

ELA	Math	Other Indicator	AYP Decision	Improvement Classification
Pass	Pass	Pass	Meets AYP	Okay
Fail	Pass	Pass	Fails AYP	Watch (ELA)
Pass	Fail	Pass	Fails AYP	Watch (Math)
Pass	Pass	Fail	Fails AYP	Watch (OI)
Fail	Fail	Pass	Fails AYP	Watch (ELA & Math)
Fail	Pass	Fail	Fails AYP	Watch (ELA & OI)
Pass	Fail	Fail	Fails AYP	Watch (Math & OI)
Fail	Fail	Fail	Fails AYP	Watch (All 3)

For example, a failure in year 1 in ELA followed by a failure in ELA in year 2 results in an INOI designation. However, failure in ELA in year 1 followed by a failure in Math but passage in ELA in year 2 would not result in an INOI designation. Instead, the school would move from being on watch for ELA to on watch for Math. A failure in Math in year 3 would move the school into INOI status (See Tables 10 & 11).

Table 11. Two Year Designation Table

Year 1 AYP Classification	Year 2 AYP Classification	2-Year Achievement Designation*
Meets AYP	Meets AYP	Okay
Fails AYP	Meets AYP	Okay
Meets AYP	Fails AYP (any combo)	Watch (“combo” elements)
Fails AYP (ELA)	Fails AYP (Math)	Watch (math)
Fails AYP (Math)	Fails AYP (ELA)	Watch (ELA)
Fails AYP (ELA)	Fails AYP (ELA)	In Need of Improvement
Fails AYP (Math)	Fails AYP (Math)	In Need of Improvement
Fails AYP (Both)	Fails AYP (either)	In Need of Improvement
Fails AYP (OI)	Meets AYP	Okay
Meets AYP	Fails AYP (OI)	Watch (OI)
Fails AYP (OI)	Fails AYP (OI)	In Need of Improvement

* Does not include an exhaustive set of designation possibilities and should be considered illustrative

Once identified as INOI, a school must successfully pass the associated AYP criteria for two consecutive years to have the label removed. Following similar logic, a school designated as INOI because of Math failure would need to pass the Math requirements for two consecutive years to have the label removed. A failure in ELA in either of those two years, but not both consecutively, would not affect the school’s INOI status, but the school would now be watched relative to ELA performance.

Exemplary and High Achievement Designations

While the AYP classification system is designed to identify schools that require technical assistance and support, it is also designed to identify schools that are exceeding state expectations for performance. This is a key attribute of the model if it is to identify schools that can serve as models for "like" lower performing schools.

As required by state statute, the State Board of Education developed criteria to be used in designating schools as high achieving and exemplary. The following is the basic set of criteria that must be met for a school/school district to earn such a distinction.

To receive either distinction, the school or school district must have met AYP for two consecutive years. In addition to meeting AYP in the current school year, the school or school district must meet these requirements:

To be designated as High Achieving:

1) The percentage of students in each identifiable subgroup that score at or above the level of "meets standard" in each subject area must be significantly greater than the annual measurable objective or PAC requirement; ***or***
2) For the school as a whole (not subgroups), the reduction in the percentage of non-proficient students (students scoring at or above meets standard) must decrease by significantly more than 10% from the previous school year.

To be designated as Exemplary:

1) The percentage of students in each identifiable subgroup that score at or above the level of "meets standard" in each subject area must be significantly greater than the annual measurable objective or PAC requirement; ***and***
2) For the school as a whole (not subgroups), the reduction in the percentage of non-proficient students (students scoring at or above meets standard) must decrease by significantly more than 10% from the previous school year.

For the PAC comparisons, "significantly greater" is based on a one-tailed 95% confidence interval. This means that the lower tail of the observed PAC for the school and subgroup, when relevant, must be greater than the annual measurable objective or PAC in the content domain. For the reduction in non-proficiency comparison, "significantly more" is based on a one-tailed 75% confidence interval meaning that the lower tail of the observed decrease for the whole school must be greater than 10%.

For example, the PAC objective in ELA at the elementary level is 30%. Using the formula provided previously for calculating the standard error of the proportion and the z-score transformation, for a school of 25 students the observed PAC in ELA would have to be 46.5% or higher to be judged as significantly greater than the annual measurable objective of 30%. Likewise, using the formula for the standard error of the difference in proportions and assuming that the size of the school was not different in the previous year, the PAC rate for

the school in the previous year would have had to have been 28.7% or lower for the reduction in non-proficiency to be judged as significant.

To aid in making significance judgments, the Department of Education will publish tools available at its website that enable easy analysis of PAC and Safe Harbor comparisons. The tool ("AYP Calculator") will assist schools and school districts in recalculating AYP for the purpose of AYP appeals but can also be used to target the level of performance necessary to be considered high achieving or exemplary.

Most Improved Designations

All school and school district AYP classifications and achievement designations discussed thus far result from distinguishing between proficient and non-proficient students or students that either meet or fail to meet standard. As required by the NCLB Act, performance on the state tests used to calculate AYP is reported relative to 4 achievement levels, not just the two that affect the calculation of PAC. In addition to the "meets standard" level of achievement, students can be classified as "exceeding standard", "approaching standard", or as "developing/emergent."

In keeping with the intent of the NCLB Act, it is important that students be challenged to progress regardless of their current level of achievement. For example, a student who demonstrates through test performance that he/she is at the Meets Standard achievement level ought to be challenged to exceed expectations. Likewise, a student who demonstrates the lowest level of achievement (developing/emergent) should be encouraged and rewarded for significant progress even if his/her progress is still below the meets standard cut point but is greater than where he/she began.

Toward this end, a school or school district that demonstrates significant movement of students into higher achievement levels in both English Language Arts and Mathematics may be recognized as demonstrating significant improvement. Significant improvement is based on a reduction of at least 7.5% in the percentage of students in the lowest achievement level and an increase of at least 7.5% in the percentage of students in the highest achievement level. In addition, there cannot be an observed decrease in the percentage of proficient students for any identifiable disaggregated group.

Reporting of Annual AYP Judgments

The state reports annual judgments in multiple ways. First, the Department issues a press release that summarizes annual AYP findings. The Department does not release preliminary findings to the public, but may release information after school districts have made final designations. Detailed profiles of schools and lists of schools and associated designations are provided on the department web page (www.doe.nv.gov) following final designations. The final designations are made by August 1st.

Additionally, Senate Bill 1 requires that information relative to both AYP and INOI classifications be published annually. This includes a listing in the annual accountability

report of those schools failing AYP. Additionally, schools identified as INOI must be listed along with an indication of the number of consecutive years in which they have had that label. It is important to note that this reporting requirement is state specific and is not required by NCLB.

Section Summary

The general approach taken to formally classify schools and report findings was shared. Because school districts share significant authority in this process, school district staff should review this material in conjunction with district guidance.

The steps and processes described are applied to all Nevada public schools with few exceptions (Private schools are not governed by the described NCLB and SB1 accountability requirements). There are rare situations and special circumstances that prohibit the "easy" application of these rules to certain schools. Some exceptions are described next.

Special Circumstances

Public schools in Nevada share many characteristics, and the majority of schools share similar grade configurations. For example, most schools follow K-5, K-6, 6-8, 7-8, or 9-12 grade configurations. However, there are anomalous configurations such as 5-8 or K-12. Moreover, some schools serve only specialized subgroups of students or contain magnet programs that serve a specific group of students. These anomalies prohibit the application or make the application of the general AYP rules more difficult.

Small Schools

As noted above, a minimum n-size is not necessarily required to make statistical comparisons, but a policy decision has been made to only compare results for subgroups if and when their aggregated total is at least 25. This creates a problem when the total school population, aggregated across testing grades, is less than 25. It is important to note that these schools must still be judged in some fashion. They cannot be exempted from the AYP determination process.

To deal with this circumstance, the typical statistical steps are applied to small schools regardless of total size. Small schools are provided some extra flexibility in appealing classifications. The extra flexibility is related to the stability of findings based on very small sample sizes. For example, small schools can introduce local assessment data that might change the interpretation of overall school performance. Local assessments have to align to state standards and must be judged to be of high technical quality in order to be used in this fashion.

Additionally, the state will aggregate data for very small schools across years until the n = 25 threshold is achieved. The state will typically conduct the current year analysis based on the current enrollment figures. If findings are negative, it will aggregate results to include up to the two previous years of data in order to achieve a sample size as close to 25 as possible. The same basic aggregation rules are applied with the addition of summing across years.

Anomalous Grade Configurations

There are two general anomalies associated with grade configurations that affect the application of AYP. The first is anomalous grade spans that bridge two or more levels of instructional programming (i.e. elementary & middle, middle & high, elementary through high). For example, some schools in the state encompass grades 5 through 8. Grade 5 is typically considered as part of the elementary program, while grade 8 is considered part of the middle school program. For these schools, data was collapsed across all grades in which assessments were administered. In practice, K-8 schools were judged against elementary statewide AYP targets, grade 5-8 schools were judged using middle school statewide AYP targets, and K-12 schools were split using the elementary targets to judge performance among K-8 students and high school targets from students in grades 9-12.

The second grade configuration anomaly is the instance in which the school only serves students in grades in which no state tests are administered (i.e. grades k-2). In this instance, "other indicator" information is the primary determinant for classifying the school. In 2005-06 an attempt will be made to "backtrack" student performance to K-2 schools from the schools they feed. Backtracking will involve the results of 3rd grade students only.

Alternative Programs

NCLB makes no distinctions regarding school types and the application of AYP, and instead reinforces the need to hold all publicly enrolled students, schools, school districts accountable under the AYP system. Hence, the rules that apply to traditional schools in Nevada are applied to what have been termed "alternative schools."

The majority of students enrolled in alternative schools/programs are there as a result of difficulties they have experienced in traditional settings. The achievement levels among these students are consistently lower than achievement levels among the general population. Moreover, most of these schools serve students that are assumed to be at-risk for academic failure.

Regardless, there is no allowable exception. Alternative schools, like all traditional schools, are judged using the standard achievement, participation, and other indicators. As is true with traditional schools, school achievement is based only on students who have been enrolled in the school for a full academic year.

Schools or Special Programs for Students with Disabilities

Just as there are alternative schools/programs, in Nevada there are a handful of public schools dedicated to serving students with disabilities. Some of the dedicated programs only serve students with rare and extremely debilitating disabilities. In many Nevada public schools, there are special programs within the school that serve a particular disabled population. These programs are at times referred to as magnet programs. Often they serve students that are <u>not</u> zoned for enrollment in that location.

As with alternative schools, no exceptions can be made for students enrolled in public schools/programs. The federal law requires that they be accounted for and that the school must be judged relative to their achievement, etc.

There has been some consideration of applying the achievement scores earned by these students back to their "zoned" school. There has also been consideration of treating magnet programs within schools as a school unto themselves. Firm policy has not been developed at this juncture and until a formal change is made, these students as a subgroup are treated as all other students enrolled in the school in which the assessments are administered.

Charter Schools

There is no distinction to be made between traditional public schools and public charter schools in Nevada in terms of the application of AYP at the school level. However, Senate Bill 1, passed after the federal review of the Nevada Accountability Work Plan, prevents the aggregation of performance within charter schools up to the sponsoring school district level.

Once again, the same AYP rules apply to charter schools just as to all other public schools. The performance of students enrolled in charter schools for a full academic year is used to judge charter schools. The performance of students in charter schools is aggregated to the state level. The state is responsible for and held accountable for all students enrolled in its public educational programs.

Correctional Programs

As is true for alternative programs, special programs, and charter schools, the federal government to date has not allowed a distinction to be made between traditional public schools and correctional facilities that provide educational programs. Therefore, at this juncture, correctional programs are reviewed annually by applying the same AYP rules used with traditional educational programs.

One of the chief concerns raised regarding the students in these programs is that their typical stay in the facility is less than a full academic year. As is true with all other schools, only students that have been enrolled in the program for a full academic year are included in AYP analyses. However, the correctional programs are embedded within school districts, and, assuming that at least some of the students enrolled in the correctional programs have been enrolled in the school district for longer periods of time, the results for students may be aggregated up to the school district level even when their performance has not impacted the school level judgment.